Biostatistiques – BST1

Chapitre 13
Principes des tests statistiques de comparaison

José LABARERE, PU-PH

Arnaud Seigneurin, MCU-PH, Bastien Boussat, MCU-PH, Alexandre Bellier, AHU, Patrice François, PU-PH

Plan

- Objectifs
- Introduction
- Formulation de l’hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d’application des tests statistiques

Objectifs

Connaitre les étapes de la démarche hypothético-déductive

Formuler l’hypothèse nulle (H0) et l’hypothèse alternative (H1)

Mettre en œuvre le test statistique

Porter un jugement de signification statistique

Formuler une conclusion intelligible

Vérifier les conditions d’application des tests

Plan

- Objectifs
- **Introduction**
- Formulation de l’hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d’application des tests statistiques

Test statistique de comparaison

- Outil statistique de comparaison

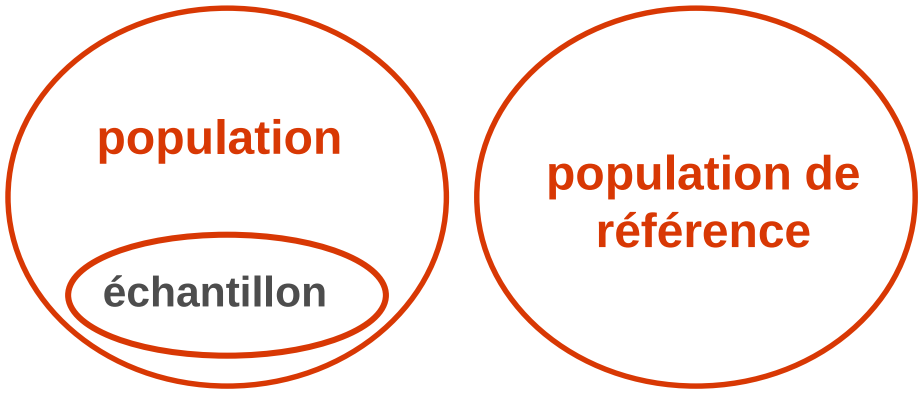
  (intervalle de confiance = outil statistique de l'estimation)

- Sert à comparer 2 ou plusieurs séries de données
  - résumées par leurs paramètres : moyenne, variance (tests paramétriques)
  - (décrites par leur distribution : tests non-paramétriques)

- Garder à l'esprit
  - On compare des paramètres estimés à partir d'échantillons issus de populations
  - La valeur de ces paramètres estimés à partir d'échantillons fluctue autour de la vraie valeur du paramètre de la population dont ils sont issus

Test statistique de comparaison

Comparer un paramètre estimé partir d'un échantillon à une valeur théorique de référence

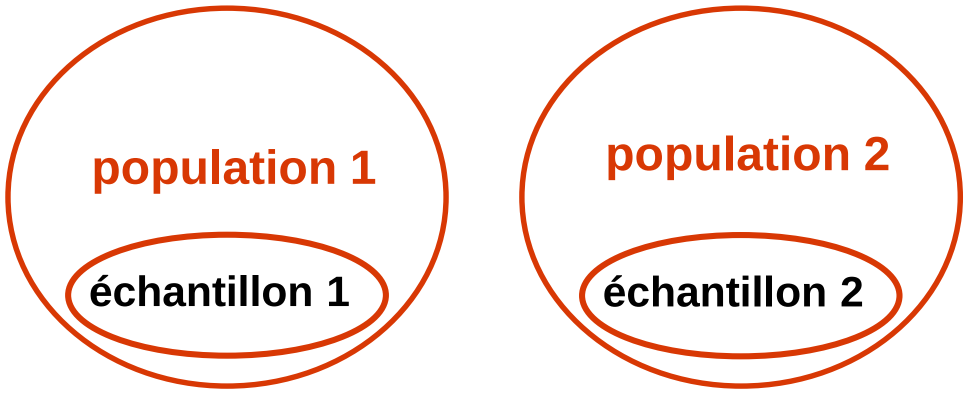


Différence observée

- Fluctuations d'échantillonnage ?
- Différence entre populations ?

Test statistique

Comparer un paramètre estimé à partir de 2 ou plusieurs échantillons



Différence observée

- Fluctuations d'échantillonnage ?
- Différence entre populations ?

Test statistique

Démarche hypothético-déductive

1. Formuler une hypothèse nulle et une hypothèse alternative
   - H0 : $\mu_{\text{population 1}} = \mu_{\text{population 2}}$
   - H1 : $\mu_{\text{population 1}} \neq \mu_{\text{population 2}}$
2. Déduire ce que devraient être les estimations si l'hypothèse nulle est vraie
   - $m_{\text{échantillon 1}} \approx m_{\text{échantillon 2}}$ aux fluctuations d'échantillonnage près
3. Confronter les estimations à ce qui était attendu sous l'hypothèse nulle
   - $m_{\text{échantillon 1}} - m_{\text{échantillon 2}}$ compatibles avec l'hypothèse nulle ?
4. Conclure
   - Non-rejet de l'hypothèse nulle ($\mu_{\text{population 1}} = \mu_{\text{population 2}}$)
   - Rejet de l'hypothèse nulle ($\mu_{\text{population 1}} = \mu_{\text{population 2}}$) → $\mu_{\text{population 1}} \neq \mu_{\text{population 2}}$

Plan

- Objectifs
- Introduction
- Formulation de l'hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d'application des tests statistiques

Formulation de l'hypothèse nulle (H0)

paramètre population 1 = paramètre population 2

$$\mu_1 = \mu_2$$

Les hypothèses sont formulées à l'aide des paramètres des populations

L'hypothèse nulle (H0) est l'hypothèse qu'on souhaite invalider (rejeter) :

- Il est plus facile de rejeter une hypothèse (un seul contre-exemple suffit)
- Valider une hypothèse demande de rechercher toutes les situations possibles et de vérifier qu'aucune d'entre-elle ne contredise cette hypothèse.

Formulation de l’hypothèse alternative (H1)

Hypothèse alternative bilatérale (H1)

paramètre population 1 ≠ **paramètre population 2**

$$\mu_1 \neq \mu_2$$

L’hypothèse alternative (H1) est l’hypothèse qui sera retenue si le test statistique rejette l’hypothèse nulle (H0)

L’hypothèse alternative est bilatérale, lorsqu’on postule que les paramètres sont différents sans chercher à déterminer le sens de cette différence.

Formulation de l’hypothèse alternative (H1)

Hypothèse alternative unilatérale (H1)

paramètre population 1 **>** **paramètre population 2**

$$(\mu_1 > \mu_2)$$

On formule une hypothèse alternative unilatérale lorsqu’on s’intéresse au sens de l’inégalité entre les 2 populations.

Le sens de l’inégalité est évident : le pourcentage de guérison est supérieur dans un groupe de patients traités par rapport à un groupe de patients non traités.

L’utilisation des tests d’hypothèse unilatéraux n’est justifiée que dans des circonstances très particulières et les revues médicales scientifiques recommandent de ne pas utiliser ces tests d’hypothèse unilatéraux pour les études de supériorité.

Plan

- Objectifs
- Introduction
- Formulation de l’hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d’application des tests statistiques

Sous l'hypothèse nulle (H0)

- **$\text{Paramètre}_{\text{population1}} = \text{Paramètre}_{\text{population2}}$**

$$\mu_1 = \mu_2$$

- **$\text{Paramètre}_{\text{échantillon1}} \approx \text{Paramètre}_{\text{échantillon2}}$**
- **$\text{Paramètre}_{\text{échantillon1}} - \text{Paramètre}_{\text{échantillon2}} \approx 0$**

$$m_1 \approx m_2 \rightarrow m_1 - m_2 \approx 0$$

En raison des fluctuations d'échantillonnage, la différence $(m_1 - m_2)$ peut prendre toutes les valeurs de R.

Mais toutes les valeurs de $(m_1 - m_2)$ n'ont pas la même probabilité.

Les valeurs de $(m_1 - m_2)$ proches de 0 sont les plus probables.

Sous l'hypothèse nulle (H0)

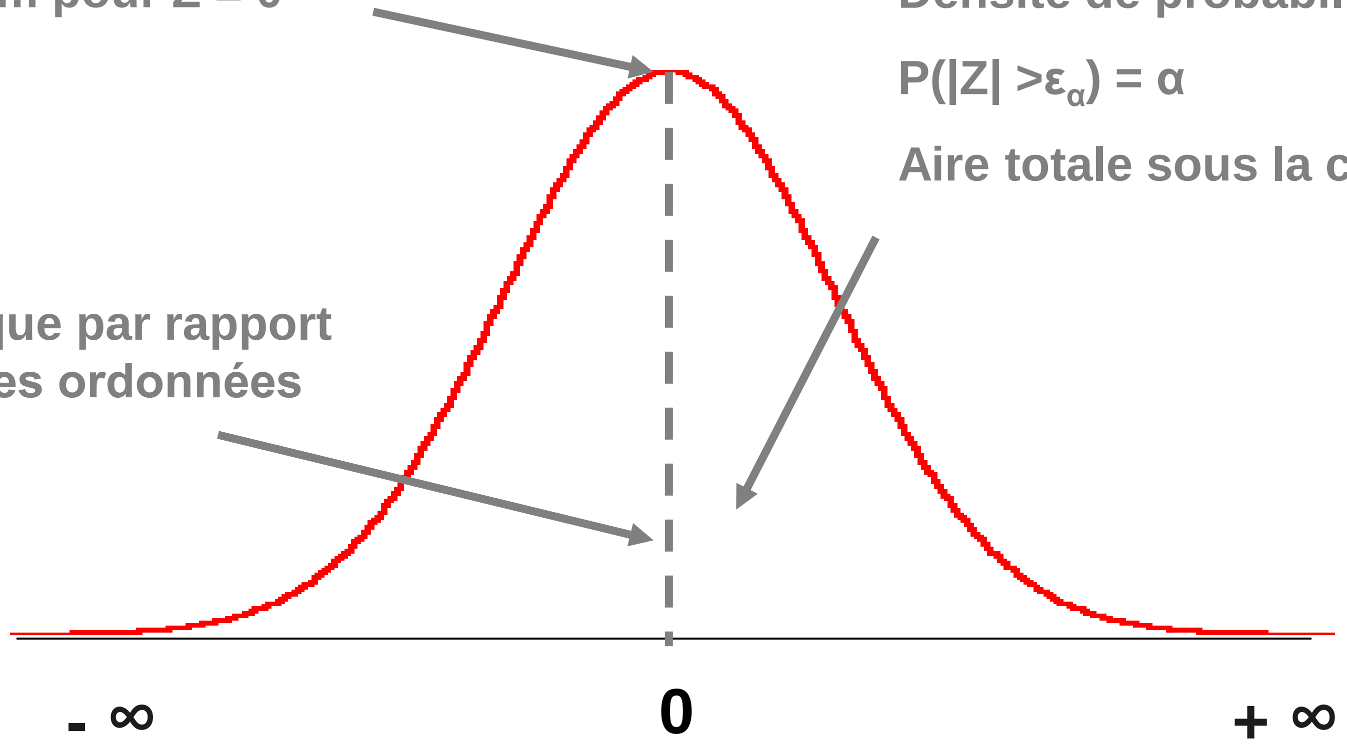
Maximum pour Z = 0

Densité de probabilité de *N*(0,1)

$P(|Z| > \varepsilon_\alpha) = \alpha$

Aire totale sous la courbe = 1 (100%)

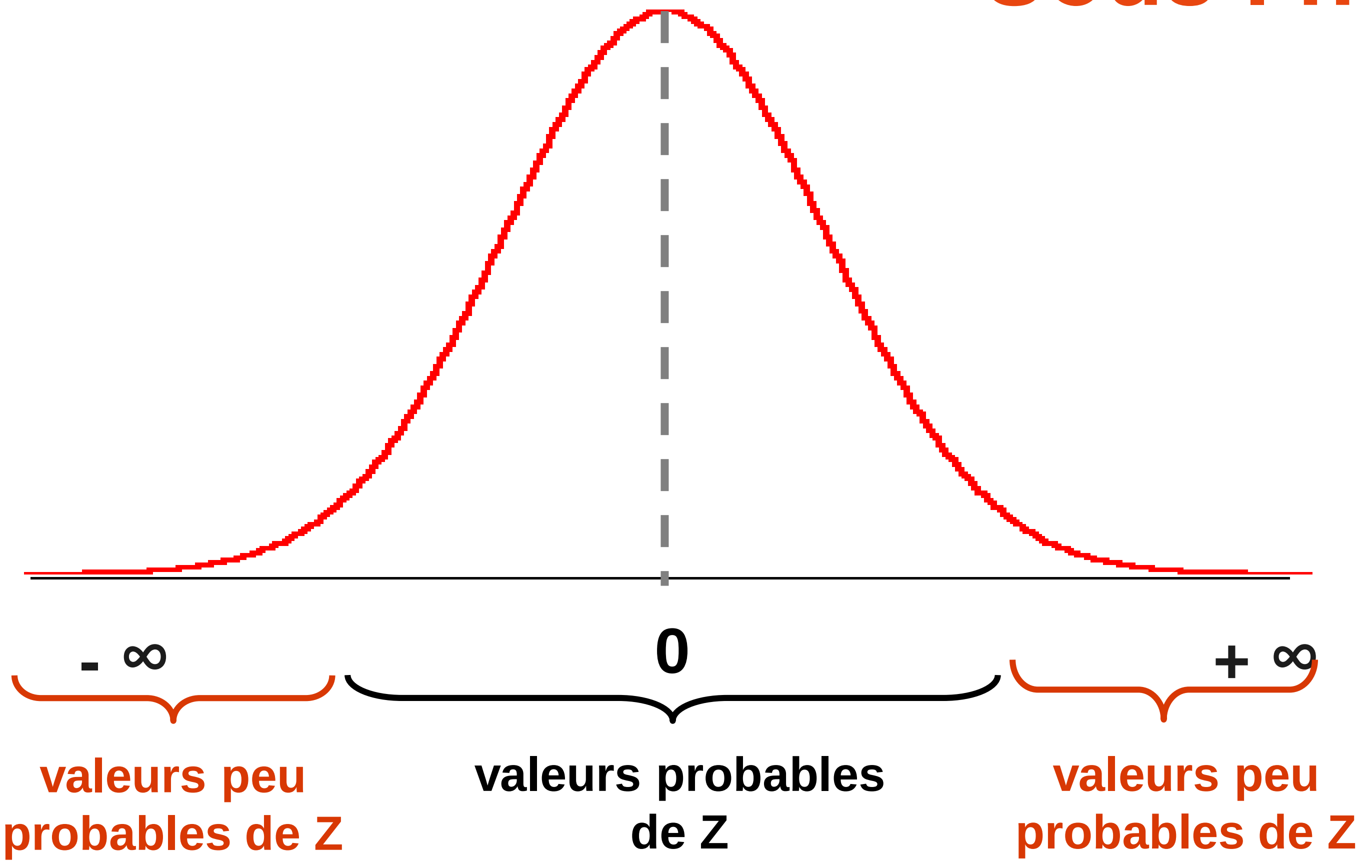
Symétrique par rapport
à l'axe des ordonnées



Abscisses : valeurs possibles de ε sous H0 ($\mu_1 = \mu_2$)

$$Z = \frac{(m_1 - m_2)}{\sqrt{\mathrm{var}(m_1 - m_2)}}$$

Sous l'hypothèse nulle (H0)



Abscisses : valeurs possibles de Z sous H0 ($\mu_1 = \mu_2$)

$$Z = \frac{(m_1 - m_2)}{\sqrt{var(m_1 - m_2)}}$$

Plan

- Objectifs
- Introduction
- Formulation de l'hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d'application des tests statistiques

Confrontation des estimations à ce qui est attendu sous H0

1. **Calculer la valeur observée de l'écart réduit Z_o à partir des estimations m_1 et m_2 sur les échantillons**

$$Z_o = \frac{(m_1 - m_2)}{\sqrt{\mathrm{var}(m_1 - m_2)}}$$

2. **Déterminer, à l'aide de la table de l'écart réduit, la probabilité que $|Z_o|$ dépasse ε sous l'hypothèse nulle (H0).**

$P(|Z_o| > \varepsilon)$ sous H0

$Z_o = Z_{observée}$

Plan

- Objectifs
- Introduction
- Formulation de l'hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d'application des tests statistiques

Seuil de signification statistique

Fixer *a priori* un seuil alpha :

- **au dessous duquel, on rejette l'hypothèse nulle H0 ($\mu_1 = \mu_2$) et on accepte l'hypothèse alternative H1 ($\mu_1 \neq \mu_2$)**

  $P(|Z_o| > \varepsilon)$ sous H0 $< \alpha \rightarrow$ rejet de H0 et acceptation de H1

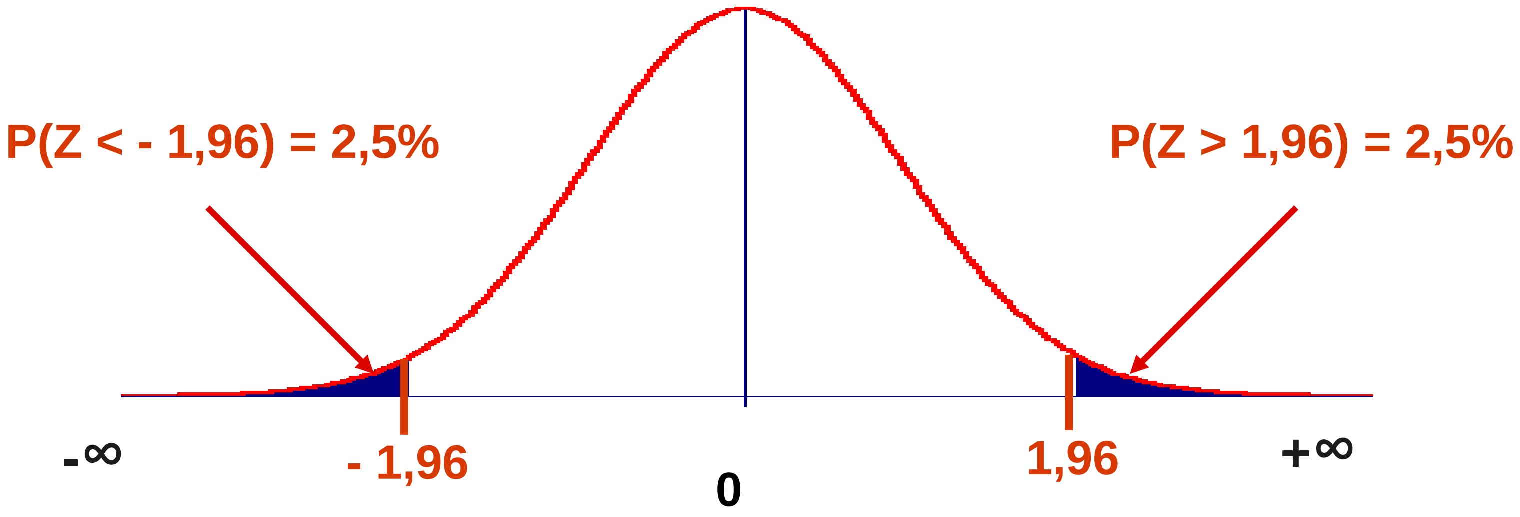
- **au dessus duquel, on ne peut pas rejeter l'hypothèse nulle H0 ($\mu_1 = \mu_2$)**

  $P(|Z_o| > \varepsilon)$ sous H0 $\geq \alpha \rightarrow$ non-rejet de H0

Classiquement : seuil de signification statistique $\alpha = 0.05$ en santé et biologie

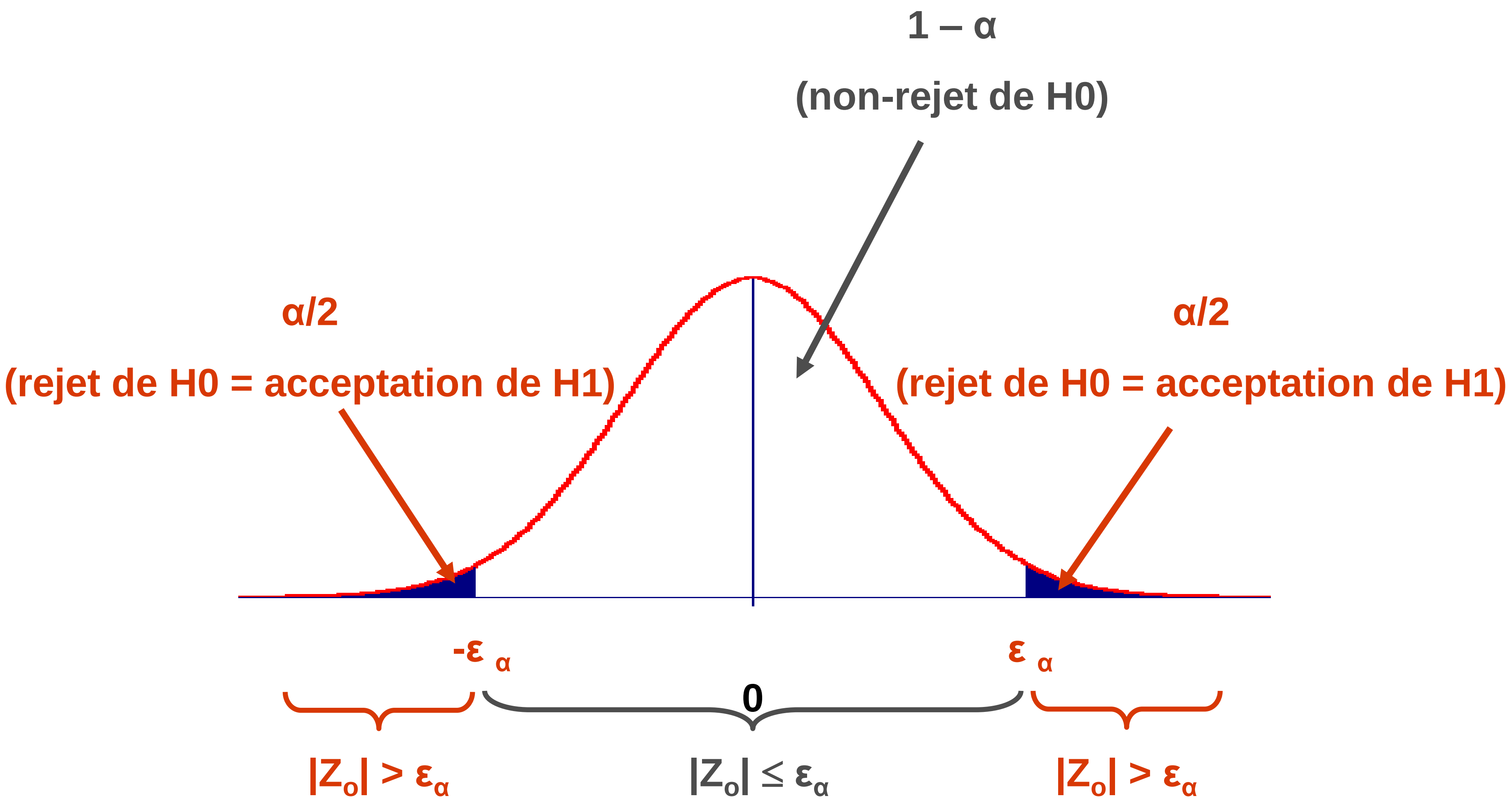
Densité de probabilité de loi normale centrée réduite $N(0,1)$

α = 5% (0,05) $\rightarrow |\varepsilon_\alpha| = 1{,}96$



Abscisses : valeurs possibles de Z sous H0 ($\mu_1 = \mu_2$)

$$Z = \frac{(m_1 - m_2)}{\sqrt{\mathrm{var}(m_1 - m_2)}}$$



Abscisses : valeurs possibles de Z sous H0 ($\mu_1 = \mu_2$)

$$Z = \frac{(m_1 - m_2)}{\sqrt{\mathrm{var}(m_1 - m_2)}}$$

Z_o : valeur observée de Z sur l'échantillon

Conclusion basée sur le test de comparaison

- **Non-rejet de H0**
  - On ne met pas en évidence de différence statistiquement significative de la valeur du paramètre comparé entre les 2 échantillons (m_1 et m_2)
  - Cela ne permet pas de conclure à l'absence de différence pour le paramètre comparé entre les 2 populations (μ_1 et μ_2)
  - « absence of evidence is not evidence of absence »

- **Rejet de H0 → acceptation de H1**
  - Le paramètre estimé diffère de manière statistiquement significative entre les 2 échantillons (m_1 et m_2)
  - On conclut que le paramètre diffère entre les 2 populations ($\mu_1 \neq \mu_2$), avec un risque d'erreur α

Plan

- Objectifs
- Introduction
- Formulation de l’hypothèse nulle et alternative
- Déduction de ce que devraient être les estimations sous H0
- Confrontation des estimations à ce qui est attendu sous H0
- Règle de décision et conclusion
- Conditions d’application des tests statistiques

Conditions d’application des tests statistiques

- Les tests statistiques sont basés sur des lois de distribution théorique (loi normale, loi de Student, loi du X^2)
- Ces lois sont strictes et nécessitent la vérification de conditions d’application :
  - Indépendance des observations
  - Spécifiques à chaque test (écart-réduit, t de Student, X^2…)

Messages clés

- Le test statistique de comparaison repose sur une démarche hypothético-déductive
- Le seuil de signification statistique est fixé a priori à 0,05 (5%) en santé et biologie
- L’indépendance des observations est une condition d’application commune à tous les tests statistiques de comparaison

Mentions légales